

ABSTRACT

A method for generating animated sequences of talking heads in text-to-speech applications wherein a processor samples a plurality of frames comprising 5 image samples. Representative parameters are extracted from the image samples and stored in an animation library. The processor also samples a plurality of multiphones comprising images together with their associated sounds. The processor extracts parameters 10 from these images comprising data characterizing mouth shapes, maps, rules, or equations, and stores the resulting parameters and sound information in a coarticulation library. The animated sequence begins with the processor considering an input phoneme 15 sequence, recalling from the coarticulation library parameters associated with that sequence, and selecting appropriate image samples from the animation library based on that sequence. The image samples are concatenated together, and the corresponding sound is 20 output, to form the animated synthesis.